p. 1

# Gaussian marginal distribution

(Important for marginalisation)

$$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} \underbrace{\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}}_{\Sigma}\right)$$

Let's define a precision matrix $\Lambda = \Sigma^{-1}$

We need to marginalise: $p(x_1) = \int p(x_1, x_2)\, dx_2$

Let's first examine the exponent of the joint

$$-\frac{1}{2}(x_1-\mu_1)^T \Lambda_{11} (x_1 - \mu_1) - \frac{1}{2}(x_1 - \mu_1)^T \Lambda_{12} (x_2 - \mu_2) -$$

$$-\frac{1}{2}(x_2-\mu_2)^T \Lambda_{21} (x_1 - \mu_1) - \frac{1}{2}(x_2 - \mu_2)^T \Lambda_{22} (x_2 - \mu_2) =$$

$$= -\frac{1}{2}\left(x_2^T \Lambda_{22} x_2 - 2 x_2^T \Lambda_{22} \left(\mu_{22} - \Lambda_{22}^{-1} \Lambda_{21} (x_1 - \mu_1)\right) -\right.$$

$$- 2 x_1^T \Lambda_{12} \mu_2 + 2 \mu_1^T \Lambda_{12} \mu_2 + \mu_2^T \Lambda_{22} \mu_{22} +$$

$$\left. + x_1^T \Lambda_{11} x_1 - 2 x_1^T \Lambda_{11} \mu_1 + \mu_1^T \Lambda_{11} \mu_1 \right)$$

Since we know the marginal is Gaussian, we can "complete the square" (in the exponent terms).

p.2

Let's put terms with the seme precision together ($\Lambda_{ij}$)

$$-\frac{1}{2}(x_2-(\mu_2-\Lambda_{22}^{-1}\Lambda_{21}(x_1-\mu_1)))^T\Lambda_{22}(x_2-(\mu_2-\Lambda_{22}^{-1}\Lambda_{21}(x_1-\mu_1)))$$

I $$\boxed{+\frac{1}{2}(x_1^T\Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}x_1-2x_1^T\Lambda_{12}\Lambda_{22}^{-1}\Lambda_{12}\mu_1+ + \mu_1^T\Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}\mu_1) -}$$

II $$\boxed{-\frac{1}{2}(x_1^T\Lambda_{11}x_1-2x_1^T\Lambda_{11}\mu_1+\mu_1^T\Lambda_{11}\mu_1)}$$

~~Now we are going to "complete the square"~~

I $$\frac{1}{2}(x_1^T\Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}x_1-2x_1^T\Lambda_{12}\Lambda_{22}^{-1}\Lambda_{12}\mu_1+\mu_1^T\Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}\mu_1)=$$

$$\underset{\{\Lambda_{12}=\Lambda_{21}\}}{=}\frac{1}{2}((x_1-\mu_1)^T(\Lambda_{12}\Lambda_{22}^{-1}\Lambda_{12})(x_1-\mu_1))$$

II $$\frac{1}{2}(x_1^T\Lambda_{11}x_1-2x_1^T\Lambda_{11}\mu_1+\mu_1^T\Lambda_{11}\mu_1)=$$

$$=\frac{1}{2}((x_1-\mu_1)^T\Lambda_{11}(x_1-\mu_1))$$

Putting everything together now:

$$-\frac{1}{2}\left(x_2-(\mu_2-\Lambda_{22}^{-1}\Lambda_{21}(x_1-\mu_1))\right)^T\Lambda_{22}\left(x_2-(\mu_2-\Lambda_{22}^{-1}\Lambda_{21}(x_1-\mu_1))\right)-$$

$$-\frac{1}{2}(x_1-\mu_1)(\Lambda_{11}-\Lambda_{12}\Lambda_{22}^{-1}\Lambda_{12})(x_1-\mu_1)$$

We have two exponents:

$$p(x_1,x_2)=\frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma|^{\frac{1}{2}}}\exp(E_1)\exp(E_2)$$

$$E_1=-\frac{1}{2}\left(x_2-(\mu_2-\Lambda_{22}^{-1}\Lambda_{21}(x_1-\mu_1))\right)^T\Lambda_{22}\left(x_2-(\mu_2-\Lambda_{22}^{-1}\Lambda_{21}(x_1-\mu_1))\right)$$

$$E_2=-\frac{1}{2}(x_1-\mu_1)(\Lambda_{11}-\Lambda_{12}\Lambda_{22}^{-1}\Lambda_{12})(x_1-\mu_1)$$

$$p(x_1)=\int p(x_1,x_2)\,dx_2=\frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma|^{\frac{1}{2}}}\int\exp(E_1)\exp(E_2)\,dx_2$$

Since $E_2$ is independent of $x_2$,

$$p(x_1)=\frac{\exp(E_2)}{(2\pi)^{\frac{D}{2}}|\Sigma|^{\frac{1}{2}}}\int\exp(E_1)\,dx_2$$

It must integrate to 1, i.e. $\frac{1}{Z}\int\exp(E_1)\,dx=1$,

so we have a distribution of the following type

$$\frac{1}{(2\pi)^{\frac{d}{2}}|S|^{\frac{1}{2}}}\int e^{-\frac{1}{2}(y-\mu_y)^T S^{-1}(y-\mu_y)}\,dy=1$$

Again, let's look at $E_1$

$$\frac{1}{Z}\int \exp(E_1)\, dx_2 = 1$$

~~...~~ $$\frac{1}{Z} = \frac{1}{(2\pi)^{\frac{D_2}{2}} |S_2|^{\frac{1}{2}}}$$

$\frac{D_2}{2}$ because we integrate over $x_2$

$|S_2|$ → covariance corresponding to $E_1$

Let's look at $E_1$:

$$S_2 = \Lambda_{22}^{-1}$$

So, $$\int \exp(E_1)\, dx_2 = (2\pi)^{\frac{D_2}{2}} |\Lambda_{22}^{-1}|^{\frac{1}{2}}$$

Then, $$p(x_1) = \underbrace{(2\pi)^{\frac{D_2}{2}} |\Lambda_{22}^{-1}|^{\frac{1}{2}} \cdot \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}}}_{\frac{1}{(2\pi)^{\frac{D-D_2}{2}} |\Lambda_{22}^{-1}|^{-\frac{1}{2}} |\Sigma|^{\frac{1}{2}}}} \cdot \exp(E_2)$$

(5)

Now, what is the determinant

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A|\,|D - CA^{-1}B|$$

So, $|\Sigma| = |\Sigma_{11}|\,|\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}|$

$$\Lambda_{22}^{-1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \quad \text{(Schur complement)}$$

$$|\Lambda_{22}^{-1}|^{-\frac{1}{2}}\,|\Sigma|^{\frac{1}{2}} = |\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}|^{-\frac{1}{2}}\,|\Sigma_{11}|^{\frac{1}{2}}\,|\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}|^{\frac{1}{2}} =$$

$$= |\Sigma_{11}|^{\frac{1}{2}}$$

$$p(x_1) = \frac{1}{(2\pi)^{\frac{D_1}{2}}\,|\Sigma_{11}|^{\frac{1}{2}}} \cdot e^{\overbrace{-\frac{1}{2}(x_1-\mu_1)^T(\Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21})(x_1-\mu_1)}^{E_2}}$$

$D_1 = D - D_2$ (dimensionality of $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is the sum of $\dim(x_1) + \dim(x_2)$)

Since $\Sigma_{11}^{-1} = \Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}$,

$$p(x_1) = \frac{1}{(2\pi)^{\frac{D_1}{2}}\,|\Sigma_{11}|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(x_1-\mu_1)\Sigma_{11}^{-1}(x_1-\mu_1)} =$$

$$= \underline{\mathcal{N}(\mu_1, \Sigma_{11})}\,!$$